Self-Managed Platforms

And the Data Center of the Future

Leading information technology companies are working on adaptive, Self-Managed, self-optimizing systems that will help enterprise customers meet their biggest IT business challenges: spiraling costs, security issues, and the need for greater applications throughput.

This white paper summarizes today's problems, describes several industry initiatives, and discusses Intel's research efforts in this space. Centered on the concept of proactive building blocks, Intel® technologies will enable computer industry vendors to create dynamic solutions that will improve data centers of the future.

Contents

Introduction

Somewhere, deep in the server farm, a cooling fan has failed. First one, then a number of blade CPUs start to heat up. An intelligent, adaptive IT system quickly detects an airflow problem that could compromise the entire rack. The system moves all processing from the impaired blades to other, healthier servers. It then shuts down the affected equipment before climbing temperatures have a chance to damage the hardware. For users who were running applications on the compromised server, it's as if nothing had happened – their processing never missed a beat.

This is a simple example of what is being variously called adaptive, or self-managed systems. With more built-in "intelligence" than today's solutions, these systems can react more quickly and efficiently to anomalies in normal operation.

Major computer industry players have launched initiatives that are intended to transform the data center by creating the next generation of systems management technologies. Intel Corporation is one of them, devoting considerable resources to researching what the company has termed "self-managed platforms."

Intel research and development is investigating potential solutions at a fundamental, building block level. These solutions – taking a cross-platform, systems view – will allow Intel customers and partners to create Self-Managed products and services essential for the data center of the future. Specifically, Intel is developing the concept of proactive building blocks: modules that combine hardware, firmware, and software to provide low-level visibility and platform control required by adaptive systems. (We describe these building blocks in more detail later.)

The Problems Research is Working to Solve

Adaptive systems have come into the limelight for a number of compelling reasons. These reasons include several particularly intractable problems CIOs must address.

For example: A primary charter of today's IT executives is to cut costs. Special emphasis falls on finding ways to lower total cost of ownership (TCO). However, systems complexity and operational costs continue their upward spiral (despite the best efforts of CIOs and IT Managers).

One of the reasons that overall data center and infrastructure costs continue to rise while reductions to TCO remains an elusive goal, is that humans continue to be involved in mundane, day-to-day IT tasks. This leaves little room for focusing on improvements in business processes, adopting new technology, and optimizing resources. As a result, labor costs continue to be a significant part of IT expense.

Giga Information Group, the Boston-based research house, reports that labor represents 46 percent of IT budgets. Even more revealing, according to Gartner Dataquest, IT workers cause 40 percent of major service disruptions. So, while on the one hand, manual labor is the solution, it also contributes to the problem. Today's corporate computing environment requires too much intervention from people to perform even the simplest and most repetitive tasks associated with maintaining and tuning the infrastructure.

Intel's IT Global Engineering Group took a close look at where IT employees spend most of their time. Here are the tasks that take most of the time and why they're so time-intensive.

Protection from malicious attacks – Virus, worms, and distributed denial of services (DDOS) attacks are on the rise from the outside. Perimeter firewalls are ineffective for stopping attacks from the inside. In addition, mobile clients, such as PDAs and laptops, can become infected while outside the firewall and can introduce worms and viruses into the system.

Asset management – Identifying distributed computing assets in large, geographically-dispersed organizations is extremely difficult. In addition to inventory control and provisioning, locating infected systems during an attack is time-consuming and error prone.

Controlling and debugging systems – Current agent-based systems provide valuable management capabilities for security, but are easily bypassed or removed by the user. It needs an adaptive system that automatically probes and troubleshoots system errors.

Visibility – Administrators do not have the level of visibility into network traffic needed to meet application quality of service (QoS) levels or distinguish normal traffic from virus, worm, or denial-of-service attacks.

The increasing use of distributed systems and networks is creating a rapid growth in IT infrastructure complexity. Combine this complexity with the current need for human operator intervention, and it becomes increasingly difficult to lower TCO and increase security.

The Research Efforts Underway

Over the past decade there has been ongoing work to solve these and related data center TCO problems. Intel, for example, teamed up with the Distributed Management Task Force (DMTF) to create the Wired for Management (WfM) initiative. The initiative looked at a number of factors, including power consumption, software installation, software inventorying, and system health capabilities.

One key finding was that a lack of out-of-band (OOB) communications capability was a major barrier in lowering TCO. At the time, OOB capabilities were complex and costly. Today, many solutions still rely on a centralized approach that may not scale with infrastructure demands – operators use centralized consoles to spot problems and respond after the fact. By the time they are ready to act, the damage has been done. Obviously, a more proactive approach is needed.

Working Toward Self-managing Platforms

Intel and other industry leaders are building on previous work, but with a different orientation – minimizing human intervention. A key idea centers on including intelligence on the platform for a more proactive approach to managing clients, servers, or elements of the network infrastructure.

Currently, in order to better manage the infrastructure, statistical data about the state of the platform or network is gathered and analyzed automatically. Then, if certain pre-determined thresholds are exceeded, an alert is sent

to a system administrator. Remember the failed fan? The system recognizes that CPU temperature has crossed a pre-defined threshold, and alerts a human operator that there is a problem that needs attention.

The next step, also currently in deployment, is a policy-based management approach that relies on a set of rules. When these rules are violated, the human administrator is notified. For example, a network alert might require that the firewall be configured to filter out certain Simple Object Access Protocol (SOAP) operations over TCP/IP. Or, based on the type or amount of traffic, settings on network switches and routers will have to be changed to optimize packet flow. This approach, although somewhat effective, is still labor-intensive.

The industry is now moving to tie monitoring and configuration into a single system, essentially taking the administrator out of the loop. Systems will no longer be dumb devices that just send out alerts when problems crop up. Self-Managed systems could automate handling many of today's mundane data center and networking tasks. They insert intelligence into the system itself, relying on local stochastic analysis to diagnose problems and execute procedural code to solve the problem or provide a work-around. For example, in the case of the hot CPU, the administrator no longer receives a dozen alerts that something has gone wrong. Instead a single alert arrives that contains a diagnosis of the problem and, in many cases, announces that the system has automatically taken corrective action.

These systems are moving toward what IBM calls "autonomic computing." In an article[1] in the *IBM Systems Journal*, the authors defined four major characteristics of autonomic computing systems. They are:

Self-configuring – Able to adapt automatically to dynamically changing environments, allowing new features, software and servers to be added to the enterprise infrastructure with no disruption of services.

Self-managing – Able to discover, diagnose and react to disruptions by detecting and isolating the failed component, then either fixing it or introducing a replacement component without any apparent application disruption.

Self-optimizing – Able to monitor and tune resources automatically.

Self-protecting – Able to anticipate, detect, identify, and protect themselves from any attacks.

These systems will be difficult to build. They rely on the development of innovative technology, while still leveraging legacy systems and infrastructure. Intel research into Self-Managed platforms aims to provide customers with a secure technology foundation at the platform level that will help them expedite the development and deployment of products for the data center of the future. The Intel® platform provides crucial local visibility, control, and intelligence to automate the mundane tasks that are driving up IT expenditures. Figure 1 shows how these three foundation technologies could fit into client and server platform architectures.

Starting with the Foundational Technologies

Intel's research focuses on ways to provide foundational technologies that will allow computer and networking industry companies to build adaptive, self-managed platforms for the data center of the future. The goal is to create building blocks that tackle today's IT problems, and address issues likely to arise in the adaptive, self-managing, self-optimizing data centers of tomorrow.

The practical solutions Intel is working on deploy at the lower end of the stack. These proactive building blocks, implemented in hardware, firmware and software, aim to provide tamper resistant, protected space on every Intel platform. Self-Managed, self-provisioning, and self-optimizing systems on the platform could work at a local level to provide visibility into the system and help apply solutions.

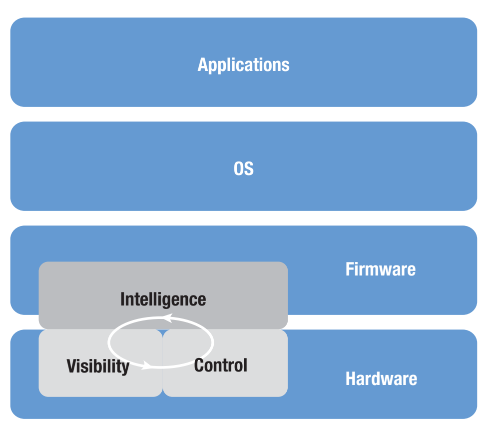

Figure 1: Intel® building blocks for self-managed platforms.

1 Ganek, A.G. and Corbi, T.A., "The Dawning of the Autonomic Computing Era," *IBM Systems Journal,* vol. 42, no. 1, 2003.

An intelligent platform at the processor level, for instance, provides excellent visibility into the moment-to-moment state of the platform and the network. It also enables rapid problem-solving on the local level. For example, intelligence in the platform could continually probe the network to determine where resources are available. If congestion suddenly occurs, an intelligent platform could instantly reroute traffic around the bottleneck.

Proactive Building Blocks

The proliferation of Intel® architecture across enterprise computing environments (from cell phones to PDAs to mobile/desktop clients to servers) provides an opportunity for a common management solution across platforms. Providers of high-level adaptive stacks could provide management solutions that interface with Intel® building blocks across the entire enterprise, from the data center to the edge of the network.

Depending on the infrastructure, Intel building blocks could be proactive themselves, combining hardware, firmware and software that operate even if the platform's operating system is down, uninstalled or otherwise unavailable. By operating autonomously within the IT ecosystem, proactive building blocks would allow out of band (OOB) communications in all system states and enable management instrumentation, alerting, configuration, and remote control.

Proactive building blocks could feature:

Persistence and tamper resistance during all states – Neither system errors nor malicious users would be able to remove the building blocks. This means that upper layer management applications could rely on complete accessibility to every network client at all times. Current software-only solutions do not provide this capability.

Out-of-band communication – The building blocks would provide a guaranteed baseline of functionality and system access. Intel customers and partners would be able to build this capability into their data center of the future product offerings. (See Figure 2.)

Industry standard specifications – Open specifications, such as XML and CIM, would be used throughout.

These innovations would help drive the architecture for self-managed platforms, bringing new benefits for enterprise manageability. These benefits would include visibility, control and intelligence. Visibility into platform state provides local data about platform operation. Capability modules inside the intelligence engine consume the platform state data and execute analysis routines that search for out-of-profile behavior. This intelligence engine may then instruct "Control" routines to change platform behavior in an automated fashion. (See Figure 3.)

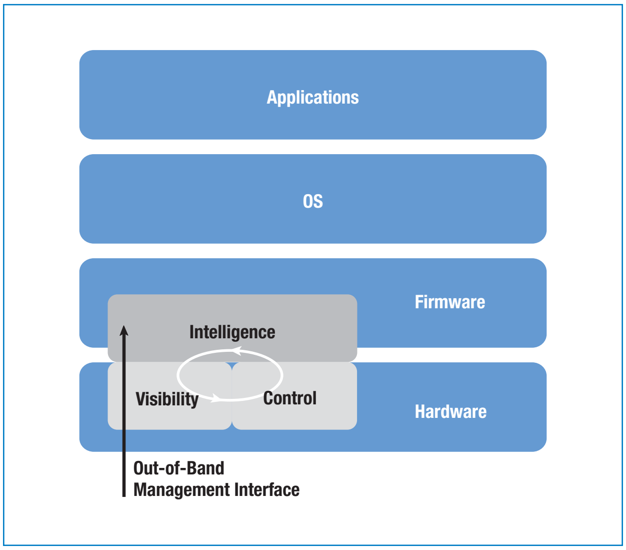

Figure 2

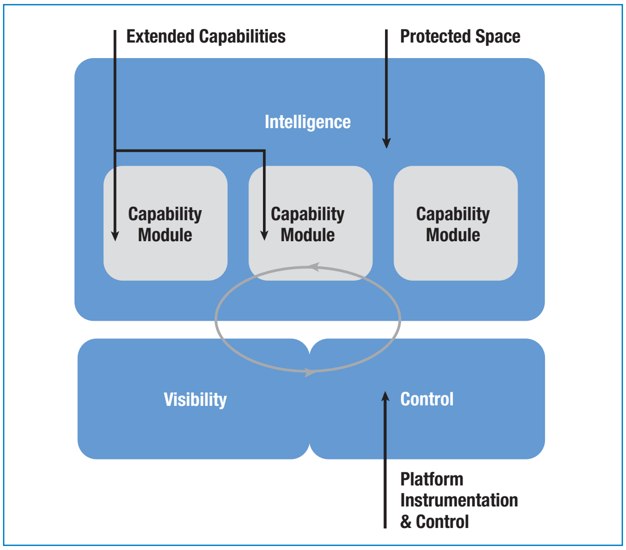

Figure 3

Working together, visibility, control and intelligence “close the loop” and enable platform automation. Let’s look at the role each would play.

Visibility

Regardless of the OS state, the building blocks can monitor network traffic and system status, detect and report anomalies. Operating at a local level, a building block can determine the nature of the problem, where it exists in the infrastructure, and whether the CPU or network is involved.

For example, a set of building blocks with a comprehensive view of network traffic detects a worm attack that triggers a sudden change in the behavior of one or more servers. It performs a stochastic analysis in real time, determines that the behavior is out of profile, and flags the problem. At the same time, it traces the connection back to other machines on the network and determines that these machines are showing anomalies as well. Obviously there is a problem. The adaptive, Self-Managed system needs to stop the worm in its tracks before it takes down the entire data center. At this point, control comes in.

Control

In the case of a virus, worm or denial-of-service attack, a building block’s visibility capabilities can isolate the signature of the attack (e.g. a SQL slammer) and correlate the signature across the data center’s platforms. It determines that other servers are failing and a common pattern quickly emerges.

The building block’s control features immediately isolate the infected servers and prevents the suspicious traffic from accessing other data center servers. Later, at their leisure, systems administrators can put a security patch in place, but in the meantime the attack has been stopped without taking the data center off line.

In another typical situation, unexpected network congestion may be causing problems. Dedicated building blocks have been constantly probing the network to determine available resources on a real-time basis. Responding to the congestion, they trigger the building block’s self-optimizing capabilities to reroute traffic and avoid the current bottleneck.

Building blocks can also handle problems originating in the CPU. For example, if the CPU has to deal with a spike in network traffic, a building block might back off a SCSI disk controller that has been sending multiple interrupts to ensure the system is operating at peak performance.

Intelligence

In order to make adaptive, self-managed, self-optimizing platforms a reality, computer industry vendors and ISVs will need intelligence at a fundamental level on a platform they can trust. Working at this level, proactive building blocks can collect and correlate low-level (firmware) statistics and send alerts using standard management protocols.

As compared to an operating system with its millions of interdependent lines of code, the building block approach provides a constrained code base that can be scrutinized line-by-line for any problems or security holes. Their OOB communications and tamper-resistant features provide a protected space where industry vendors can add value to their data center products and make their offerings more attractive to their enterprise customers.

Solving the IT Infrastructure Problem

Through its research into adaptive, self-managed platforms and the development of the proactive building block concept, Intel is addressing key issues that directly impact the development of self-managed platforms. The impact of many of the solutions discussed here remains five to seven years out. The company is taking a comprehensive, cross-platform, systems view in order to develop solutions at a fundamental level. In turn, Intel’s customers will be able to use these building block solutions to create new technologies and platforms that directly address IT’s key problems and pave the way for the data center of the future.

For More Information

Addressing IT Challenges with Self-Healing Technology
Preston Hunt and Dylan Larson
http://developer.intel.com/update/contents/it10032.htm

Computer System, Heal Thyself
Linda Dailey Paulson, *IEEE Computer*, August 2002
Available to IEEE members at:
http://www.computer.org/computer/

Distributed Management Task Force (DMTF)
Information about Common Information
Model (CIM) Standards
http://www.dmtf.org/standards/standard_cim.php

HP/Adaptive Enterprise
http://h71028.www7.hp.com/enterprise/
html/6842-0-0-0-121.html

IBM/Autonomic Computing
http://www.research.ibm.com/autonomic

Intel/Business Enterprise
Information about "Managing Complexity" and other
topics useful for business customers.
http://www.intel.com/business/bss/infrastructure/
enterprise/dsm.htm

Intel TCO Community of Interest
https://azure.intel.com/tco/index.asp

Microsoft/Dynamic Systems Initiative
http://www.microsoft.com/presspass/press/2003/
mar03/03-18dynamicsystemspr.asp

Wired for Management (WfM)
http://www.intel.com/labs/manage/wfm/index.htm